

Blickfixationstiefe in stereoskopischen VR-Umgebungen: Eine vergleichende Studie

Thies Pfeiffer Matthias Donner Marc E. Latoschik Ipke Wachsmuth

AG Wissensbasierte Systeme, Technische Fakultät, Universität Bielefeld

Abstract: Für die Mensch-Maschine-Interaktion ist die Erfassung der Aufmerksamkeit des Benutzers von großem Interesse. Für Anwendungen in der Virtuellen Realität (VR) gilt dies insbesondere, nicht zuletzt dann, wenn Virtuelle Agenten als Benutzerschnittstelle eingesetzt werden. Aktuelle Ansätze zur Bestimmung der visuellen Aufmerksamkeit verwenden meist monokulare Eyetracker und daher auch nur zweidimensionale bedeutungstragende Blickfixationen relativ zu einer Projektionsebene. Für typische Stereoskopie-basierte VR Anwendungen ist aber eine zusätzliche Berücksichtigung der Fixationstiefe notwendig, um so den Tiefenparameter für die Interaktion nutzbar zu machen, etwa für eine höhere Genauigkeit bei der Objektauswahl (Picking). Das in diesem Beitrag vorgestellte Experiment zeigt, dass bereits mit einem einfacheren binokularen Gerät leichter zwischen sich teilweise verdeckenden Objekten unterschieden werden kann. Trotz des positiven Ergebnisses kann jedoch noch keine uneingeschränkte Verbesserung der Selektionsleistung gezeigt werden. Der Beitrag schließt mit einer Diskussion weiterer Schritte mit dem Ziel, die vorgestellte Technik weiter zu verbessern.

Stichworte: Mensch-Maschine-Interaktion, Eyetracking, Virtuelle Realität

1 Einleitung

In der technisch gestützten Kommunikation, wie sie z.B. im EU-Projekt PASION (Psychologically Augmented Social Interaction Over Networks) [BMWD06] untersucht wird, spielen Augenbewegungen eine große Rolle, denn das Herstellen eines Blickkontaktes wird in Gesprächen unter anderem zur Verständigungssicherung oder zum Turn-Taking eingesetzt. Speziell für Virtuellen Agenten, wie z.B. Max [KJLW03], sind daher Verfahren zur Blickbestimmung von besonderem Interesse, um etwa ein möglichst genaues situatives Partnermodell des jeweiligen Gegenübers aufbauen zu können (Projekt A1, *Partnermodellierung*, Sonderforschungsbereich 673 *Alignment in Communication*).

Dreidimensionale Fixationen sind darüber hinaus auch für die empirische Forschung interessant. In der Psycholinguistik wird unter dem Namen *Visual World* Paradigma [TSKES95] seit einigen Jahren die Interaktion zwischen visuellem Kontext und Sprachverarbeitung untersucht. Die Tiefe des Raumes wurde dort jedoch nur indirekt berücksichtigt, aber gerade z.B. bei der Erforschung von räumlichen Präpositionen (*davor* vs. *dahinter* [GHW93]) ist dies unumgänglich. Ein zuverlässiges automatisches Verfahren zur Bestimmung der Tiefe einer Fixation eröffnet hier entscheidende neue Möglichkeiten für die Grundlagenforschung.

In der VR machen insbesondere stereoskopische Anwendungen eine Tiefenbestimmung not-

wendig, da die Tiefenwahrnehmung hier einen zentralen Stellenwert, etwa bei der Erzeugung eines hohen Immersionsfaktors, einnimmt.

Vor diesem Hintergrund werden die folgenden Fragen in diesem Beitrag bearbeitet:

1. *Welche Möglichkeiten bestehen zur Bestimmung der Fixationstiefe?*
2. *Welche Algorithmen eignen sich für eine Verwendung in der VR?*
3. *Wie gut eignen sich verschiedene Eyetracking-Geräte?*
4. *Führt die Berücksichtigung der Fixationstiefe zu Verbesserungen in der Anwendung, exemplarisch getestet an der visuellen Selektion?*

Im Abschnitt *Stand der Forschung* wird Frage 1 geklärt und die optischen Grundlagen und entsprechende Algorithmen werden vorgestellt. Anschließend wird in Abschnitt 3 ein Experiment vorgestellt, welche die Fragen 2, 3 und 4 bearbeitet. Bei der Bearbeitung von Frage 3 wurde explizit ein Vertreter aus der High-End-Klasse mit einem deutlich günstigeren Gerät verglichen, um herauszufinden, ob der Einstieg in die dreidimensionale Fixationsmessungen auch bereits mit der Low-Cost-Klasse möglich ist. Die Diskussion der Ergebnisse und ein Ausblick auf weiterführende Studien und Arbeiten schließen diesen Beitrag ab.

2 Stand der Forschung

In der Kommunikation mit Virtuellen Agenten wird Eyetracking z.B. dazu eingesetzt, Daten für natürlichere Augenbewegungen für die Virtuellen Agenten zu ermitteln. Duchowski et al. [DCC⁺04] übertragen die Augenbewegungen eines Benutzers auf einen virtuellen Avatar und belegen die Vorteile einer sichtbaren Blickrichtung für die Selektion. Wissen über die visuelle Aufmerksamkeit ist auch technisch von Interesse, wie z.B. Lübke et al. [LHNW00] zeigen, die Visualisierungsprozesse auf Bereiche im Fokus optimieren.

Besonders interessant ist die Tiefenbestimmung von Fixationen für die direkte Interaktion bzw. Manipulation in der Virtuellen Realität. Tanriverdi und Jacob [TJ00] belegen einen signifikanten Geschwindigkeitsvorteil für eine blickgesteuerte Selektion von Objekten gegenüber manueller gestischer Selektion. Ihr Selektionsverfahren kombiniert das Picking von SGI Performer mit einem Histogramm-basierten Ansatz, bei dem die Häufigkeit der Fixationen auf ein Objekt innerhalb eines Zeitfensters eine Selektion auslösen kann. Dabei wird die Blickposition eines Auges auf einer zweidimensionalen Ebene erfasst und als Ausgangspunkt für das Picking verwendet.

Vergleichbare Ansätze, die auf einem virtuellen Sehstrahl beruhen, werden von Duchowski et al. [DMC⁺02] oder Barabas et al. [BGA⁺04] verfolgt. Der Sehstrahl beginnt dabei an der Position des Auges oder des Kopfes und verläuft durch eine Fixation auf einer zweidimensionalen Ebene (in der Regel der Projektionsebene). Die Bestimmung der Tiefe einer Fixation ist bei einer solchen Vorgehensweise jedoch mit mehreren Problemen behaftet: sie ist (a) nur möglich, wenn der Sehstrahl auf eine Geometrie trifft; es besteht (b) eine Ambiguität, falls mehrere Geometrien den

Sehstrahl schneiden und es wird (c) nicht berücksichtigt, welches Auge gerade die dominante Fixationsleistung bringt. Die Probleme (a) und (b) sind in dieser Form auch für das Zeigen/Picking bekannt und es existieren verschiedene Methoden, die Erkennungsrate zu verbessern, z.B. können durch eine Berücksichtigung des Abstandes von Objekten zum Sehstrahl auch knapp verfehlte Objekte noch als fixiert angenommen werden.

Mit einem binokularen Eyetracker könnte die Konvergenz beider Augen bestimmt und die Tiefe der Fixation berechnet werden. Im Folgenden werden dazu die optischen Grundlagen rekapituliert, um im Anschluss zwei entsprechende Algorithmen vorzustellen.

2.1 Räumliches Sehen

Obwohl auf der Netzhaut des Auges nur eine zweidimensionale Projektion der Umwelt abgebildet wird, ist der Mensch in der Lage, daraus eine räumliche Vorstellung seiner Umgebung zu generieren. In der Literatur (z.B. [Gol02]) werden dazu mehrere Kriterien beschrieben:

monokulare Tiefenkriterien wie *Verdeckung*, *relative Größe/Höhe im Blickfeld*, *gewohnte Größe von Objekten*, *atmosphärische und lineare Perspektive*, der *Texturgradient* oder die *Bewegungssparallaxe* vermitteln über ein einzelnes Auge einen räumlichen Eindruck

binokulare *Querdisparität* (Unterschiede im Abbild auf der Netzhaut durch die versetzte Anordnung der Augen), *Konvergenz* (siehe Abb. 1) oder *Akkommodation* benötigen den Einsatz beider Augen

Wenn nun die Tiefe einer Fixation abgeleitet werden soll, dann stehen nur solche Kriterien zur Auswahl, die sich über ein geeignetes Verfahren messen lassen. Da die meisten der Kriterien keine sensormotorische Kopplung verlangen, kommen von den aufgezählten Kriterien nur noch die *Konvergenz* und die *Akkommodation* in Betracht, da beide in ihrer Stärke in Abhängigkeit zur Entfernung des betrachteten Objekts variieren. Mit binokularen Eyetrackern stehen Geräte zur Verfügung, mit denen die *Konvergenz* bestimmt werden kann.

2.2 Geometrische Fixationsbestimmung über Kreuzung der Blickachsen

Ein mit beiden Augen fixierter Punkt lässt sich mittels einer Geradengleichungen berechnen (Abb. 1). Gegeben sind die Positionen der beiden Augen a_{links} und a_{rechts} relativ zur Projektionsebene, sowie die über den Eyetracker ermittelten scheinbaren Fixationen s_{links} und s_{rechts} auf der Projektionsebene. Damit lassen sich folgende Gleichungen aufstellen:

$$g_{links} = a_{links} + \mu \cdot w_{links} \qquad g_{rechts} = a_{rechts} + \eta \cdot w_{rechts} \qquad (1)$$

$$w_{links} = s_{links} - a_{links} \qquad w_{rechts} = s_{rechts} - a_{rechts} \qquad (2)$$

Die Punkte f_{links} bzw. f_{rechts} auf beiden Blickachsen, die den geringsten Abstand zur jeweils anderen Gerade haben, bestimmt man als Lösung eines entsprechenden Minimierungsproblems. Der Fixationspunkt f ergibt sich dann als Mittelwert von f_{links} und f_{rechts} .

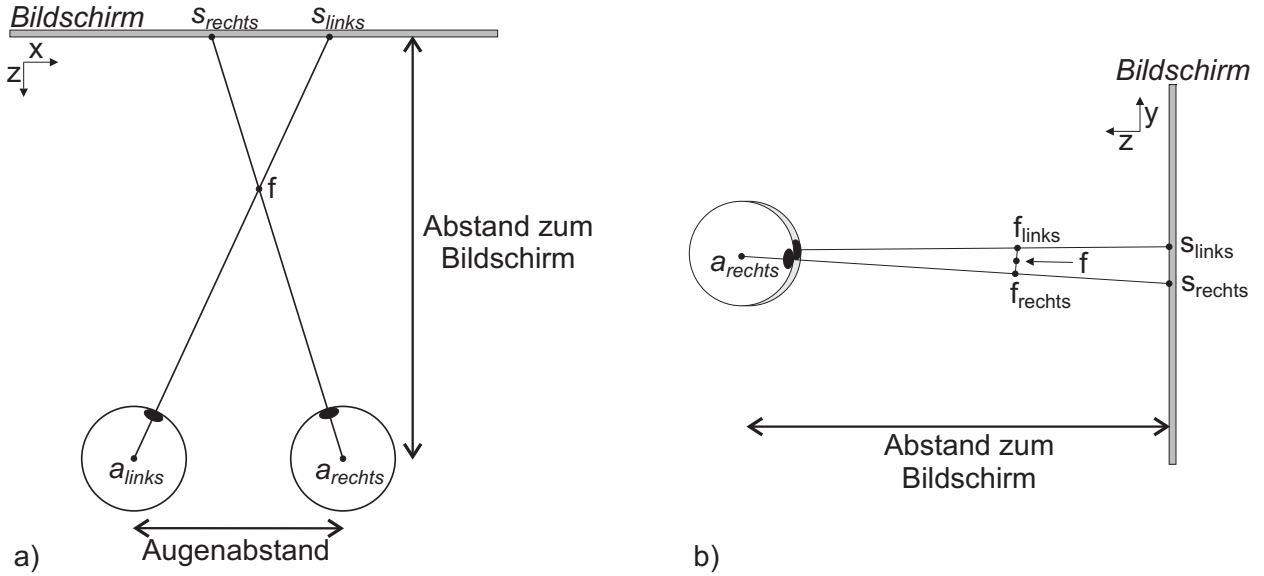


Abbildung 1: Geometrische Bestimmung der Fixationstiefe. a) Aufsicht: Die Sehstrahlen der Augen kreuzen sich in Punkt f . Die Seitenansicht b) zeigt jedoch, dass sie sich nicht scheiden.

Dieses Verfahren ist jedoch mit einigen Nachteilen behaftet. Die physischen Parameter wie Größe und Augenabstand variieren zwischen Personen und müssen stets neu vermessen werden. Darüber hinaus lässt sich der Abstand zur Projektionsfläche nicht immer genau einhalten. Essig et al. [EPR06] haben gezeigt, dass ein adaptives Verfahren unter diesen Bedingungen besser geeignet sein kann. Ihr Ansatz wird im Folgenden kurz dargestellt und im Experiment auf seine Anwendbarkeit in der VR evaluiert.

2.3 Fixationsbestimmung mittels Parametrized Self-Organizing Map

Die Parametrized Self-Organizing Map (kurz PSOM) ist eine glatte hochdimensionale Merkmalskarte [Rit93]. Sie besteht, analog zur SOM [Koh90], aus Neuronen $a \in A$ die jeweils einen eigenen Referenzvektor w_a besitzen, der eine Projektion in den Eingaberaum $X \subseteq \mathbb{R}^d$ darstellt. Für die Fixationsbestimmung ist $w_a = (x_l, y_l, x_r, y_r, x_{div})$, wobei x_l/y_l und x_r/y_r die vom Eyetracker gemessenen scheinbaren Fixationen auf der Projektionsebene sind. Da der horizontale Abstand der Fixationen entscheidend für die Bestimmung der Tiefeninformationen ist, wird er als $x_{div} = x_r - x_l$ zusätzlich in w_a aufgenommen.

Zum Training der PSOM werden nacheinander alle 27 Punkte eines dreidimensionalen $3 \times 3 \times 3$ Kalibrierungswürfels präsentiert und die entsprechenden w_a gemessen. Damit lässt sich nun eine Funktion $w(s)$ konstruieren, welche die Koordinaten des dreidimensionalen Gitters auf die Referenzvektoren abbildet. Dazu wird $w(s)$ parametrisiert zu

$$w(s) = \sum_{a \in A} H(a, s) \cdot w_a \quad (3)$$

$$\text{mit } H(a, s) = 1 \quad \text{für } s = a \quad (4)$$

$$H(a, s) = 0 \quad \forall s \neq a \text{ mit } s, a \in A \quad (5)$$

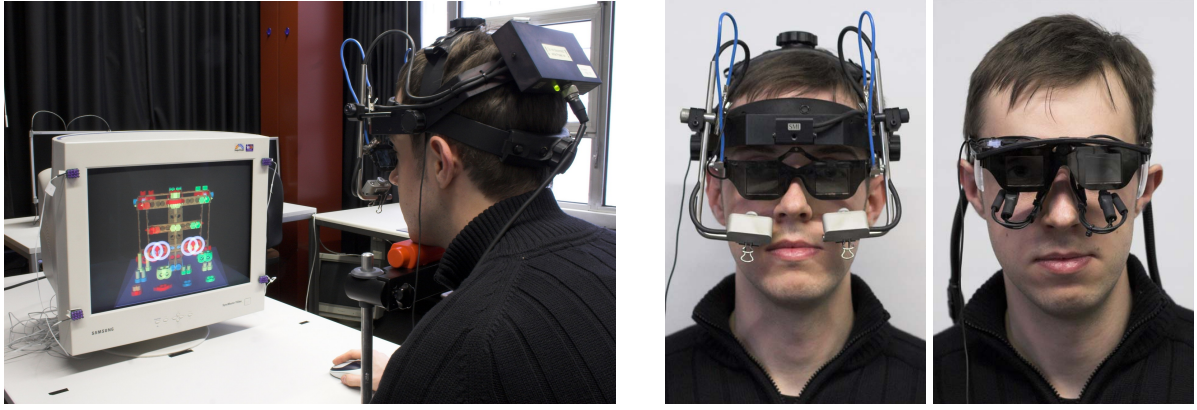


Abbildung 2: Versuchsaufbau (links) und die Eyetracker mit Shutterbrille von SMI (EyeLink I, mittig) und Arrington Research (PC60, rechts).

Im konkreten Fall ist A ein $3 \times 3 \times 3$ Gitter mit 27 Neuronen

$$A = \{a_{xyz} | a_{xyz} = x\vec{e}_x + y\vec{e}_y + z\vec{e}_z; \quad x, y, z \in \{0, 1, 2\}\} \quad (6)$$

$$H : A \times \mathbb{R}^3 \rightarrow \mathbb{R} \quad (7)$$

Um die gestellte Bedingung zu erfüllen, wird nun H mittels Produktansatz zerlegt:

$$H(x\vec{e}_x + y\vec{e}_y + z\vec{e}_z, s_x\vec{e}_x + s_y\vec{e}_y + s_z\vec{e}_z) = H^{(1)}(x, s_x) \cdot H^{(1)}(y, s_y) \cdot H^{(1)}(z, s_z) \quad (8)$$

Für die eindimensionalen Funktionen $H^{(1)} : \{0, 1, 2\} \times \mathbb{R} \rightarrow \mathbb{R}$ muss ebenso gelten:

$$H^{(1)}(n, s) = 1 \quad \text{für } s = n \quad (9)$$

$$H^{(1)}(n, s) = 0 \quad \forall s \neq n \text{ mit } s \in \mathbb{R}, n \in \{0, 1, 2\} \quad (10)$$

Da n nur drei Werte annehmen kann, lassen sich drei kubische Polynome finden, die diesen Anforderungen genügen:

$$H^{(1)}(0, s) = \frac{1}{2}s^2 - \frac{3}{2}s + 1 \quad (11)$$

$$H^{(1)}(1, s) = -s^2 + 2s \quad (12)$$

$$H^{(1)}(2, s) = \frac{1}{2}s^2 - \frac{1}{2}s \quad (13)$$

Damit ist $w(s)$ so konstruiert, dass sie die Koordinaten des dreidimensionalen Gitters auf die zweidimensionalen Augenpositionen abbilden kann. Für die Fixationsbestimmung ist nun nur noch auf numerischem Weg, über ein Gradientenabstiegsverfahren, die inverse Funktion zu bilden.

3 Ein Experiment zur Bestimmung von 3D Fixationen

Aus den Fragestellungen der Einleitung wurden die folgenden Hypothesen als Grundlage für das Experiment abgeleitet:

H1: PSOM ist genauer als geometrische Fixationsbestimmung Die Adaptivität der PSOM sollte deutliche Vorteile gegenüber der geometrischen Lösung liefern, da für letztere nicht alle notwendigen Parameter mit der notwendigen Genauigkeit manuell erfasst werden können.

H2: High-End-Gerät ist genauer als Low-Cost-Modell In der Studie wurden zwei verschiedene Eyetracker-Modelle getestet (Abb. 2): der EyeLink I von SMI als Vertreter der High-End-Geräte (> 30.000 €) und das System PC60 von Arrington Research aus dem Low-Cost Bereich (< 12.000 €). Aus den technischen Daten in Tabelle 1 ist zu entnehmen, dass das Gerät von SMI deutliche Vorteile bei Geschwindigkeit und Messgenauigkeit hat.

H3: Fixationstiefe erhöht Erfolgsrate bei der Selektion Durch den Einbezug der Tiefe sollten schwierige Fälle, in denen sich Objekte teilweise verdecken, aber deutlich unterschiedliche Tiefen aufweisen, noch disambiguiert werden können. Daher sollte dieses Verfahren eine höhere Erfolgsrate bei Objektselektionen aufweisen als solche, die 2D Fixationen mit Pickingalgorithmen verbinden.

	Arrington PC60	SMI EyeLink I
zeitliche Auflösung (in Hz)	30 / 60	250
optische Auflösung (in Pixeln)	640 × 480 / 320 × 240	keine Angabe
Abweichung zur realen Augenposition	0,25° - 1,0° Blickwinkel	< 1,0° Blickwinkel
Messgenauigkeit	0,15° Blickwinkel	0,01° Blickwinkel
Kompensation der Kopfbewegung	nicht möglich	±30° horizontal, ±20° vertikal

Tabelle 1: Technische Daten der beiden untersuchten Eyetracker.

3.1 Szenario: Stereoskopische Darstellung mittels Shutterbrille

Um die äußeren Faktoren für den Vergleich der Verfahren möglichst konstant zu halten, wurde der Abstand vom Kopf zur Projektionsebene durch eine Kinnstütze konstant auf 65cm gehalten. Die Höhe wurde jeweils so eingestellt, dass sich die Augen auf der Höhe der Oberkante des virtuellen Kalibrierungswürfels befanden (Abb. 3).

Das Experiment bestand aus zwei Durchläufen, je einem pro getesteten Eyetracker. Die Präsentation der Stimuli erfolgte mittels Shutterbrille auf einem Röhrenmonitor, da beide Eyetracking-Systeme für diesen Einsatz vorbereitet sind. Da die Kameras der Eyetracker nicht durch die Shutterbrille filmen können, wurden sie darunter angebracht.

Nach der Standardkalibrierung in 2D mit der mitgelieferten Eyetracking-Software, wurden den Teilnehmenden des Experimentes nacheinander die Punkte des Kalibrierungswürfels gezeigt. Da sich herausstellte, dass jede Person einen individuellen Zeitraum brauchte, bis die dreidimensionale Wahrnehmung vollständig einsetzte, wurde die Kalibrierung selbständig gesteuert.

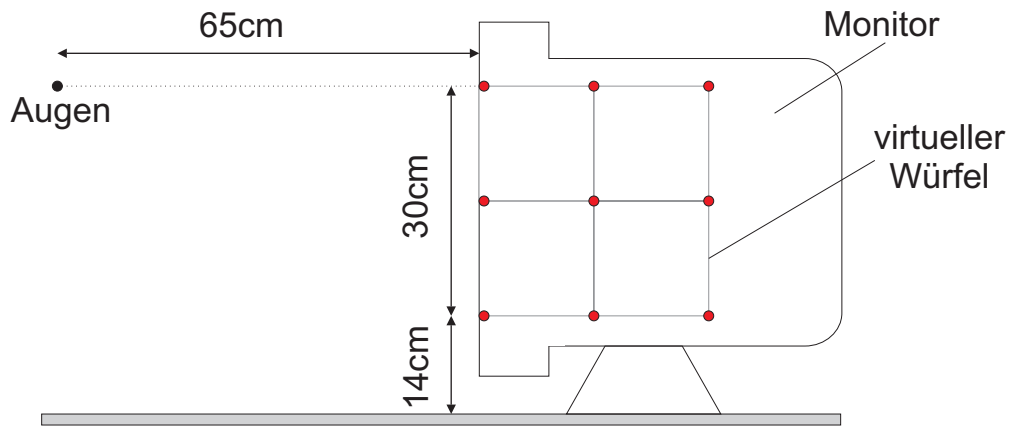


Abbildung 3: Skizze des Settings: der untersuchte virtuelle Raum erstreckt sich bis 30cm hinter die Projektionsebene.

Während des Experimentes wurde dann ein größengetreues Computermodell eines Baufixmodells gezeigt (Abb. 2). Vom Versuchsleiter wurden einzelne Bestandteile des Modells verbal beschrieben, die mit Blicken identifiziert werden sollten. Sobald dies erfolgt war, bestätigten die Probanden dies mittels Tastendruck. Intern wurden dann jeweils die 3D Fixationspunkte über beide Verfahren berechnet und alle Daten abgespeichert. Insgesamt wurden pro Person 22 Objekte auf diese Weise getestet.

3.2 Ergebnisse

Für dieses Experiment wurden insgesamt zehn Personen (vier Frauen, sechs Männer) ausgewählt. Das mittlere Alter lag bei 26,2 Jahren. Die jüngste Person war 21 Jahre alt, die älteste 41 Jahre. Alle Personen hatten normale oder korrigierte Sicht, letzteres über Kontaktlinsen. Als Kommentar gaben vier Personen an, dass sie es als schwierig erachteten, die virtuellen Kalibrierungskreuze mittels Augenbewegungen zur Überlagerung zu bekommen, und einen dreidimensionalen Eindruck zu erhalten.

Die relative Abweichung der berechneten Fixationen von den tatsächlichen Objektpositionen (hier wurde der Mittelpunkt angenommen) über alle Personen zeigen die Bagplots für die Koordinatenachsen Y und Z (Tiefe) in Abbildung 4.

Der Kolmogorow-Smirnow-Test zeigt, dass beide Messreihen nicht normalverteilt sind. Daher wurde der Mann-Whitney-Wilcoxon-Test verwendet um zu überprüfen, ob sich die Absolutwerte der Mittelwerte der beiden Messreihen signifikant unterscheiden, und ob die Mittelwerte signifikant vom Sollwert abweichen. In allen Tests wird auf ein 95-prozentiges Signifikanzniveau getestet (siehe Tabelle 2).

Bei den Messreihen beider Eyetracker bezüglich der Z-Koordinate zeigt sich, dass die Mittelwerte, die mit der PSOM berechnet wurden, signifikant näher an den Sollwerten liegen, als die Mittelwerte, die mit dem geometrischen Verfahren ermittelt wurden. Alle Mittelwerte weichen jedoch signifikant vom Sollwert ab. Die Mittelwerte der Messreihen des Arrington Eyetrackers sind näher am Sollwert als die Mittelwerte der Messreihen des SMI Eyetrackers. Die Standardabwei-

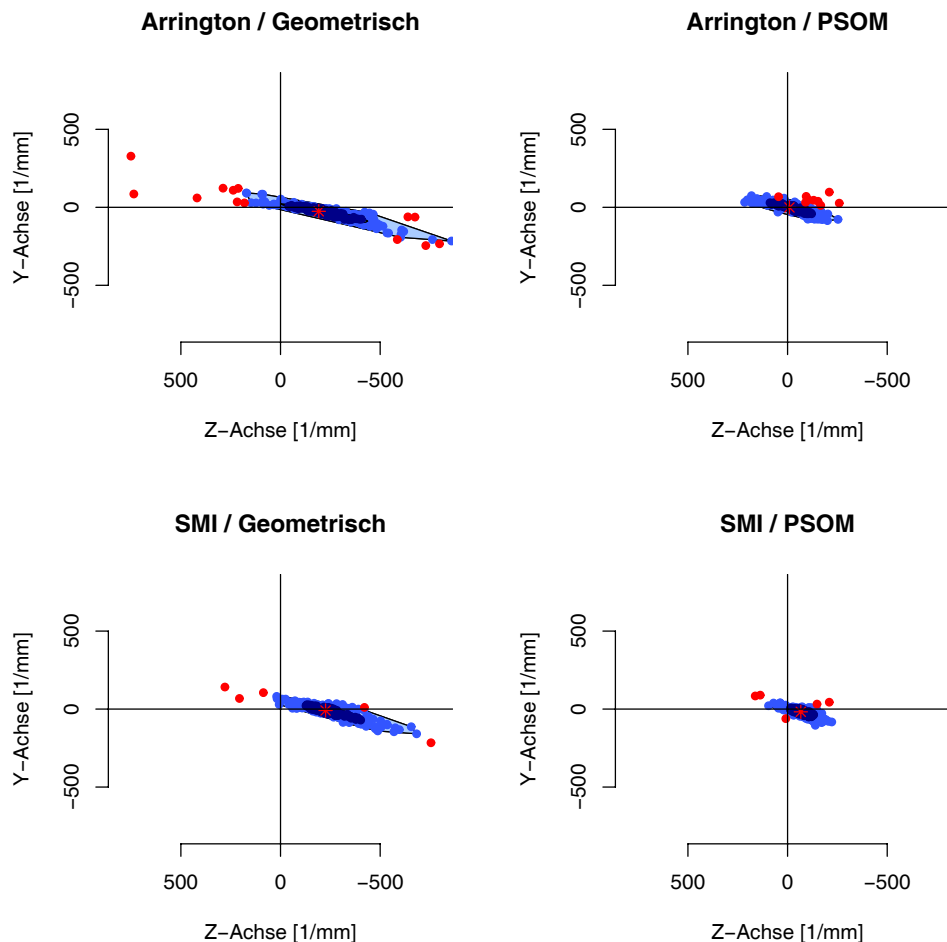


Abbildung 4: *Koordinatenabweichungen*. Bagplots der Verteilung der gemessenen Werte relativ zum Sollwert. Die dunkle Fläche enthält Fixationen mit weniger als 50%, die hellere mit weniger als 75% Abweichung. Der Stern stellt jeweils den Mittelwert da.

chungen des SMI Eyetrackers sind jedoch geringer als die des Arringtons. Bei beiden Eyetrackern lässt sich beobachten, dass die Standardabweichung bei der PSOM kleiner als bei dem geometrischen Verfahren ist.

Neben der Erfassung der quantitativen Genauigkeit wurden auch die qualitativen Auswirkungen an einem Anwendungsszenario getestet: der Objektselektion. Verglichen wurde, ob über eine dreidimensionale Fixationsbestimmung mehr Selektionen korrekt zugeordnet werden können, als über eine zweidimensionale. Im Zweidimensionalen wurde der euklidische Abstand der von der Eyetracking-Software gelieferten Bildschirmkoordinaten für beide Augen und den Bildschirmkoordinaten aller 22 Objekte berechnet. Das Objekt, dessen Projektionsmittelpunkt den geringsten Abstand von einer der beiden gemessenen Fixationen aufwies, wurde als selektiert gewertet und mit dem vorgegebenen Sollwert verglichen. Auf dreidimensionale Fixationen lässt sich diese Vorgehensweise analog übertragen. Von den 22 Objekten bildeten vier einen für die Disambiguierung kritischen Bereich (siehe Problem b), da sie hintereinander angeordnet waren und sich damit ihre Projektionen überlappten.

Messreihe	Normalverteilt	Mittelwert	Mittelwerts- unterschied	Abweichung Sollwert	Standard- abweichung
Arr. geom.	nein, $p < 0,001$	-195,77 mm	signifikant $p < 0,001$	sig. $p < 0,001$	526,69 mm
Arr. PSOM	ja, $p = 0,943$	-18,75 mm		sig. $p = 0,005$	96,92 mm
SMI geom.	nein, $p = 0,038$	-248,55 mm	signifikant $p < 0,001$	sig. $p < 0,001$	149,3 mm
SMI PSOM	ja, $p = 0,661$	-70,57 mm		sig. $p < 0,001$	60,06 mm

Tabelle 2: Der Vergleich der beiden Verfahren zeigt einen signifikanten Unterschied der Mittelwerte der Fixationstiefe zu Gunsten des PSOM-Verfahrens.

Von den insgesamt 220 Objektselektionen (22 pro Person) wurden mit dem zweidimensionalen Verfahren 165 (75%) korrekt erkannt. Mit dem dreidimensionalen Verfahren waren es nur 92 (42%). Im kritischen Bereich der Objekte 17 bis 20 (40 Selektionen) liegt jedoch wie erwartet das 3D Verfahren mit 17 (42%) zu 12 (30%) korrekten Selektionen vorne. Abbildung 5 zeigt die Ergebnisse aufgeschlüsselt für die jeweiligen Objekte. Die Zuordnung der Zahlen zu den Objekten im Modell ist aus Abbildung 6 ersichtlich.

4 Diskussion

Aus den Ergebnissen des Experimentes lassen sich nun die Folgerungen für die drei Hypothesen ableiten:

H1 bestätigt: PSOM genauer als geometrische Fixationsbestimmung Bei beiden

Eyetrackern liefern die von der PSOM ermittelten Fixationen im Mittel eine signifikant geringere Abweichung vom Sollwert. Auch die Standardabweichung der PSOM ist kleiner.

H2 (teilweise) widerlegt: Low-Cost Modell im Experiment genauer Obwohl die

Standardabweichungen beim EyeLink I geringer sind, sind die Mittelwerte weiter von den Sollwerten entfernt als beim PC60. Ein Grund dafür könnte sein, dass beim EyeLink I die zweidimensionale Kalibrierung, mit unter der Shutterbrille positionierten Kameras, sehr schwierig war. Häufig wurde sie von der mitgelieferten Software als „poor“ eingestuft. Somit waren die Ausgangsdaten weniger präzise. Die Aussage gilt also nur eingeschränkt für die verwendete technische Konfiguration und im Speziellen für das Projektionsverfahren über Shutterbrillen.

H3 (teilweise) widerlegt: Fixationstiefe verschlechtert Erfolgsrate Die Berücksichtigung der Fixationstiefe führt auf die ganze Szene bezogen zu deutlich weniger (42% zu 75%) korrekten Selektionen. Jedoch zeigen sich im kritischen Bereich (Objekte 17 bis 20 in Abbildung 6), in dem durch die Verdeckung der Objekte bei einer 2D-Selektion Ambiguitäten entstehen, die Vorteile der Berücksichtigung der Fixationstiefe (42% zu 30%). Ein Vergleich der vom jeweiligen 3D-Verfahren berechneten XY-Koordinaten mit den vom Eyetracker gelieferten

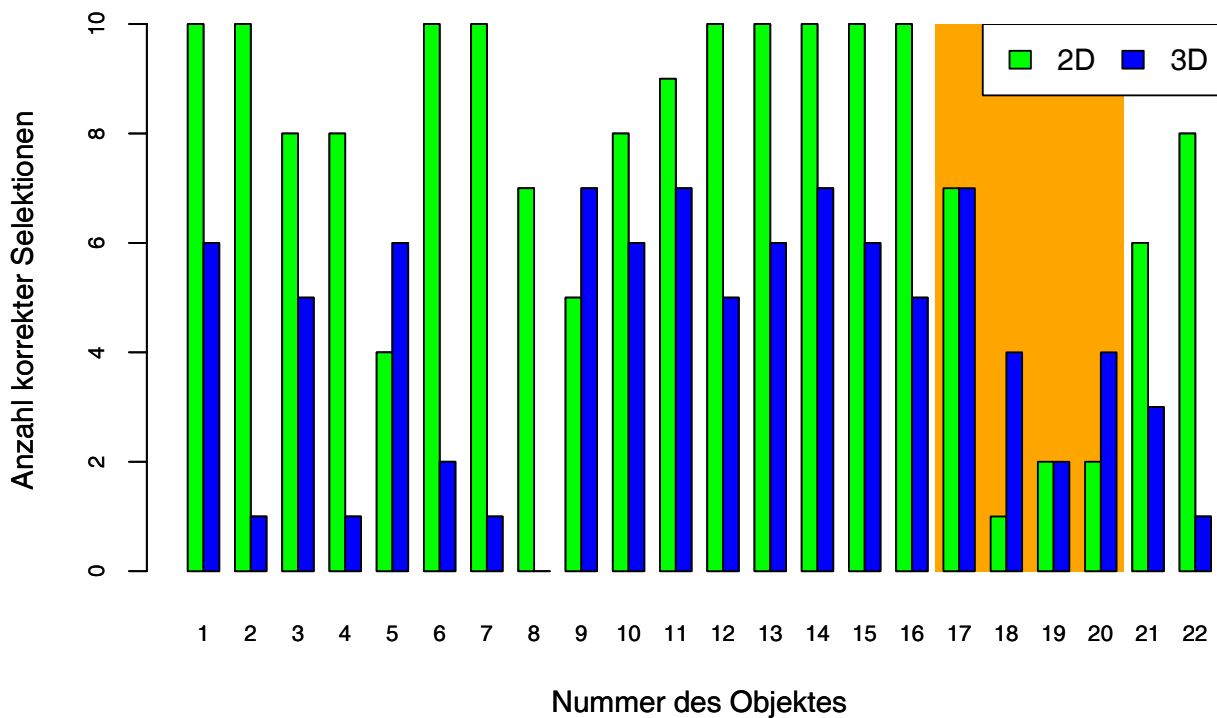


Abbildung 5: Histogramm der korrekten Selektionen über alle zehn Versuchsreihen (vergleiche Abb. 6). Der kritische Bereich der sich überlappenden Objekte (17 bis 20) ist farblich hervorgehoben.

zeigte, dass die berechneten Werte stärker streuen. Dies lässt sich durch eine ungenaue dreidimensionale Kalibrierung erklären, die auf generellen Schwierigkeiten mit der Kalibrierung unter dem Projektionsverfahren beruhen kann, auf die einige der Testpersonen hingewiesen haben. Ein weiterer Grund kann die kleine, und damit nur schwer aufzulösende, maximale Blickwinkeldifferenz von lediglich 0,9 Grad zwischen Fixationen auf der vordersten und solchen auf der hintersten Tiefenebene sein.

Die Ergebnisse zeigen, dass die Berechnung dreidimensionaler Fixationen möglich ist und die Probleme der Ambiguität im kritischen Bereich aufgelöst werden können. Im vorgestellten Szenario lässt sich das Verfahren auf Grund der Streuung der XY-Koordinaten jedoch momentan noch nicht generell empfehlen. Zu viele Parameter beeinflussen die Berechnungen. Beachtliche Teilverbesserungen können durch den Einsatz adaptiver Verfahren, wie der PSOM, erreicht werden, jedoch werden von dieser im verwendeten Setting noch nicht alle Parameter berücksichtigt. Weiterhin stören äußere Faktoren, wie die für die Wahrnehmung problematische Projektionstechnik (Ghosting) und der eingeschränkte Betrachtungsraum.

Das vorliegende Experiment ist Teil einer Reihe. In einem geplanten Vergleichsexperiment über einem analogen Szenario mit realen Objekten sollen mögliche Einflüsse der Projektionstechnik auf die natürliche Konvergenz untersucht werden. Danach wird die Genauigkeit des Eye-trackers von Arrington Research in einer immersiven VR Installation mit einer auf polarisiertem Licht basierenden Projektionstechnik untersucht. Besondere Herausforderungen sind dabei die freie Beweglichkeit des Benutzers und der auf drei Meter erweiterte Interaktionsraum. Mo-

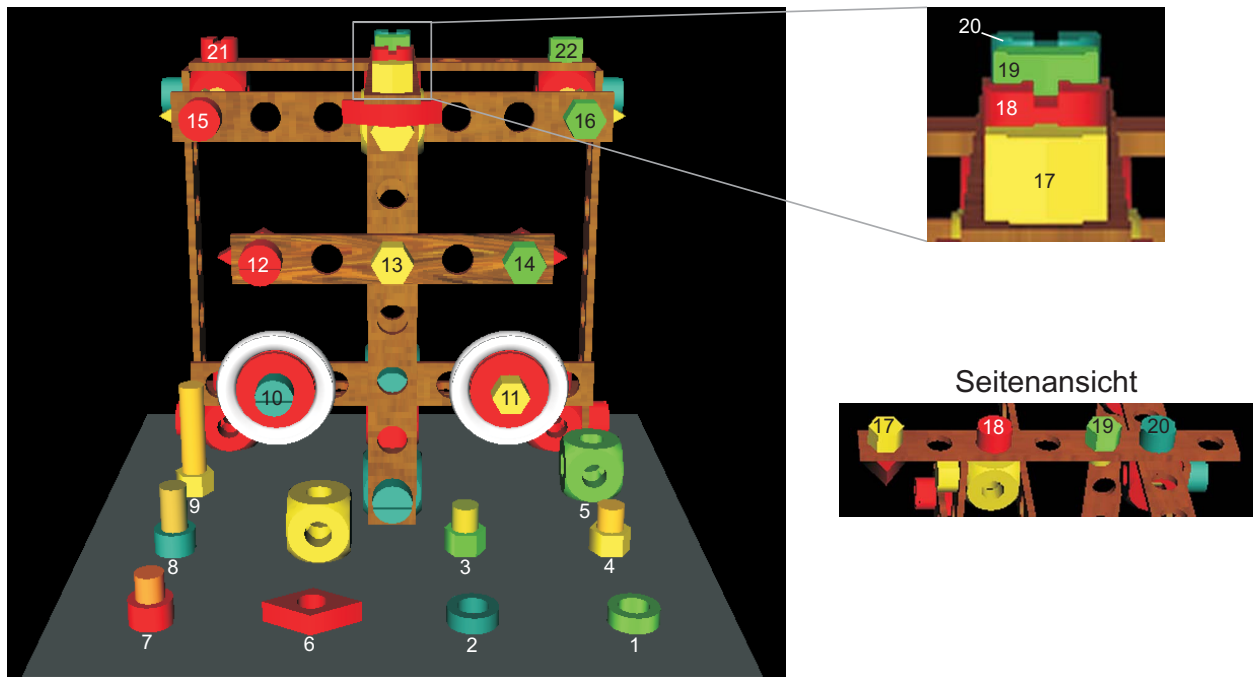


Abbildung 6: Position der Objekte im Modell (links). Die Objekte 17 bis 20 stellen für die 2D Selektion eine besondere Schwierigkeit dar (rechts).

mentan wird die Abbildung der Augenpositionen auf eine 2D Position von der mitgelieferten Eyetracking-Software erledigt. Um diesen unnötigen Zwischenschritt zu überspringen, soll das PSOM-Verfahren erweitert werden, so dass es direkt auf den Augenpositionen arbeitet.

In dicht gefüllten Umgebungen, oder zur Disambiguierung zwischen Hintergrund (Wände, etc.) und Vordergrund erscheint momentan ein hybrider Ansatz, bei dem in der Regel das 2D Verfahren eingesetzt wird und nur die Ambiguitäten über das 3D Verfahren aufgelöst werden, angebracht.

4.1 Danksagung

Diese Arbeit wurde von der Deutschen Forschungsgemeinschaft im Rahmen des Sonderforschungsbereiches 673 *Alignment in Communication* und von der EU im Rahmen des Projektes PASION (Psychologically Augmented Social Interaction Over Networks) unterstützt.

Literatur

- [BGA⁺04] BARABAS, JAMES, ROBERT B. GOLDSTEIN, HENRY APFELBAUM, RUSSELL L. WOODS, ROBERT G. GIORGI und ELI PELI: *Tracking the line of primary gaze in a walking simulator: Modeling and calibration*. Behavior Research Methods, Instruments and Computers, 36 (4):757–770, 2004.

- [BMWD06] BRUGNOLI, M. C., F. MORABITO, R. WALKER und F. DAVIDE: *The PASION Project: Psychologically Augmented Social Interaction Over Networks*. PsychNology, 4(1):103–116, 2006.
- [DCC⁺04] DUCHOWSKI, ANDREW T., NATHAN COURNIA, BRIAN CUMMING, DANIEL MCCALLUM, ANAND GRAMOPADHYE, JOEL GREENSTEIN, SAJAY SADASIVAN und RICHARD A. TYRRELL: *Visual Deictic Reference in a Collaborative Virtual Environment*. Eye Tracking Research and Applications Symposium 2004, San Antonio, 2004.
- [DMC⁺02] DUCHOWSKI, ANDREW T., ERIC MEDLIN, NATHAN COURNIA, HUNTER MURPHY, ANAND GRAMOPADHYE, SANTOSH NAIR, JEENAL VORAH und BRIAN MELLO: *3D Eye Movement Analysis*. 2002.
- [EPR06] ESSIG, KAI, MARC POMPLUN und HELGE RITTER: *A neural network for 3D gaze recording with binocular eye trackers*. The International Journal of Parallel, Emergent and Distributed Systems, 21 Nr. 2:79–95, 2006.
- [GHW93] GRABOWSKI, J., T. HERRMANN und P. WEISS: *Wenn „vor“ gleich „hinter“ ist - zur multiplen Determination des Verstehens von Richtungspräpositionen*. Kognitionswissenschaft, Nr. 3:171–183, 1993.
- [Gol02] GOLDSTEIN, E. BRUCE: *Wahrnehmungspsychologie*. Spektrum Akademischer Verlag, 2002.
- [KJLW03] KOPP, S., B. JUNG, N. LESSMANN und I. WACHSMUTH: *Max - A Multimodal Assistant in Virtual Reality Construction*. KI-Künstliche Intelligenz, 4:11–17, 2003.
- [Koh90] KOHONEN, TEUVO: *The self-organizing map*. Proceedings of IEEE, 78:1464–1480, 1990.
- [LHNW00] LUEBKE, DAVID, BENJAMIN HALLEN, DALE NEWFIELD und BENJAMIN WATSON: *Perceptually Driven Simplification Using Gaze-Directed Rendering*. Technischer Bericht, University of Virginia, 2000.
- [Rit93] RITTER, HELGE: *Parametrized self-organizing maps*. ICANN93 Proceedings, Seiten 568–577, 1993.
- [TJ00] TANRIVERDI, VILDAN und ROBERT J. K. JACOB: *Interacting with eye movements in virtual environments*. In: *Conference on Human Factors in Computing Systems, CHI 2000*, Seiten 265–272, New York, 2000. ACM Press.
- [TSKES95] TANENHAUS, M.K., M.J. SPIVEY-KNOWLTON, K.M. EBERHARD und J.C. SEDIVY: *Integration of visual and linguistic information in spoken language comprehension*. Science, 268:1632–1634, 1995.